Diffusion Model Alignment Using Direct Preference Optimization

Year: 2023 | Citations: 471 | Authors: Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou

---

Abstract

Large language models (LLMs) are fine-tuned using human comparison data with Reinforcement Learning from Human Feedback (RLHF) methods to make them better aligned with users' preferences. In contrast to LLMs, human preference learning has not been widely explored in text-to-image diffusion models; the best existing approach is to fine-tune a pretrained model using carefully curated high quality images and captions to improve visual appeal and text alignment. We propose Diffusion-DPO, a method to align diffusion models to human preferences by directly optimizing on human comparison data. Diffusion-DPO is adapted from the recently developed Direct Preference Optimization (DPO) [36], a simpler alternative to RLHF which directly optimizes a policy that best satisfies human preferences under a classification objective. We re-formulate DPO to account for a diffusion model notion of likelihood, utilizing the evidence lower bound to derive a differentiable objective. Using the Pick-a-Pic dataset of 851K crowdsourced pairwise preferences, we fine-tune the base model of the state-of-the-art Stable Diffusion XL (SDXL)-1.0 model with Diffusion-DPO. Our fine-tuned base model significantly outperforms both base SDXL-1.0 and the larger SDXL-1.0 model consisting of an additional refinement model in human evaluation, improving visual appeal and prompt alignment. We also develop a variant that uses AI feedback and has comparable performance to training on human preferences, opening the door for scaling of diffusion model alignment methods.